Teachers' Assessment of Antisocial Behavior in Kindergarten: Physical Aggression and Measurement Bias across Gender

Jantine L. Spilt[1], Helma M.Y. Koomen[1], Jochem T. Thijs[2], Reinoud D. Stoel[3], Aryan van der Leij[1]

Correspondence to: Jantine Spilt, Department of Education, University of Amsterdam, P.O. Box 94208, 1090 GE Amsterdam, the Netherlands. Phone: +31(0)205251272; Fax: +31(0)205251200; Email: J.L.Spilt@uva.nl

[1] Department of Education, University of Amsterdam, The Netherlands

[2] Department of Social Science, Utrecht University, The Netherlands

[3] Netherlands Forensic Institute, The Hague, The Netherlands

This is an Author's Accepted Manuscript of an article published in Journal of Psychoeducational Assessment, 2010, 28, 129-138. © SAGE Publications, the final version is available online at http://dx.doi.org/10.1177/0734282909340236

Abstract

A confirmatory factor analytic (CFA) study was conducted to obtain evidence for physical aggression as a distinct construct of nonaggressive antisocial behavior in young children. Second, we investigated factorial invariance across gender. Teachers completed the Preschool Behavior Questionnaire (PBQ) for two independent samples of kindergartners (N=487; N=1557). Behavior items were selected representing physically aggressive versus nonaggressive antisocial behavior. To obtain support for the two-factor model, we also examined associations with subtypes of internalizing behavior. CFA confirmed that physical aggression constitutes a distinct construct from nonaggressive antisocial behavior for young children. In support of the model, differential associations with internalizing behavior and different outcomes with respect to gender differences were found. Factorial non-invariance across gender was found for physical aggression, and explanations for those gender differences are discussed.

Keywords: physical aggression, measurement invariance, gender differences, internalizing behavior, kindergarten children

Teachers' Assessment of Antisocial Behavior in Kindergarten: Physical Aggression and Measurement Bias across Gender

Developmental scholars (e.g., Broidy et al., 2003; Tremblay, 2000) have repeatedly argued for the need of pure assessments of physical aggression not confounded by other forms of antisocial behavior, as early physical aggression is considered a major risk factor for chronic aggression and juvenile violence. However, this distinctiveness of physical aggression from other antisocial behavior has not yet been stringently tested in young children. Therefore, we sought to obtain evidence for physical aggression as a separate construct for both boys and girls using confirmatory factor analyses (CFA). To advance understanding of the nature of physical aggression that is measured in boys and girls, we tested factorial invariance across sex.

Kindergartners who are kicking, hitting, and threatening other children display more worrisome behavior than children who don't want to share toys or are inconsiderate to others. While occasional use of physical aggression (i.e., use of verbal threat and physical force to harm others or damage objects) is normative, persistent physical aggression is associated with many adjustment problems, including social adversity and school failure (Cote, Vaillancourt, LeBlanc, Nagin, & Tremblay, 2006; Joussemet, Frank Vitaro, Barker, Côté, Nagin et al., 2008; NICHD Early Child Care Research Network, 2004; Silver, Measelle, Armstrong, & Essex, 2005; Stipek & Miles, 2008). Research has consistently identified a group of children (4.4-16.6%) that follows a high and stable trajectory of physical aggression from an early age onwards, whereas most children follow low to moderate desisting trajectories (for a review see: Nagin & Tremblay, 2005). Therefore, a screening strategy is advocated focused on physical aggression rather than disruptive behavior in general to detect young at-risk children (Broidy et al., 2003; Joussemet et al., 2008; Tremblay, 2000).

This is an Author's Accepted Manuscript of an article published in Journal of Psychoeducational Assessment, 2010, 28, 129-138. © SAGE Publications, the final version is available online at http://dx.doi.org/10.1177/0734282909340236

There is considerable evidence to consider physical aggression as a distinct dimension of antisocial behavior. Physical aggression has been successfully distinguished from relational aggression in which relationships serve as means to harm others (e.g., Crick & Grotpeter, 1995; Vaillancourt, Brendgen, Boivin, & Tremblay, 2003), and nonaggressive forms of antisocial behavior (Achenbach, Conners, Quay, Verhulst, & Howell, 1989; Frick, Lahey, Loeber, Tannenbaum, Van Horn et al., 1993; Loeber & Schmaling, 1985a/b; Quay, 1987; Tackett, Krueger, Sawyer, & Graetz, 2003). This latter subtype represents behaviors that violate social norms and are potentially harmful to others but not physically or verbally abusive. The distinction between physical aggression and nonaggressive antisocial behavior is supported by differences in etiology and risk factors such as parental traits and psychopathology, parenting skills and monitoring, stability, and heritability (Eley, Lichtenstein, & Stevenson, 1999; Monuteaux, Fitzmaurice, Blacker, Buka, & Biederman, 2004; Nigg & Hinshaw, 1998; Patrick, Snyder, Schrepferman, & Snyder, 2005; Stanger, Achenbach, & Verhulst, 1997; Tremblay, 2000). In addition, internalizing behavior seems negatively linked to physical aggression (Crick & Grotpeter, 1995), but positively to nonaggressive antisocial behavior (Kazdin, 1992). Probably most imperative, however, is the finding that physical aggression is a unique predictor of juvenile violence (Cote et al., 2006; Loeber & Hay, 1997; Nagin & Tremblay, 1999; NICHD Early Child Care Research Network, 2004), whereas nonaggressive antisocial behavior seems a precursor of nonviolent delinquency in boys (Nagin & Tremblay, 1999).

Importantly, current evidence is largely restricted to boys in middle and late childhood, while research that explicitly focuses on early childhood or girls is still sparse (for a review, see Moreland & Dumas, 2008). Scholars have repeatedly notified that, in this age range, physical aggression is often confounded with other forms of antisocial behaviors (e.g., Broidy et al., 2003; Joussemet et al., 2008; Tremblay, 2000). Moreover, there have been no studies

using CFA in samples of young children testing the assumption of physical aggression as a distinct construct.

The view that manifestations of physical aggression are gender specific further underscores the need to distinguish physically aggressive acts from less gender-specific forms of antisocial behavior. There is clear evidence for boys being more physically aggressive than girls across different age ranges, while gender differences in indirect aggression (Card, Stucky, Sawalani, & Little, 2008) and nonaggressive conduct problems appear trivial (Maughan, Pickles, Rowe, Costello, & Angold, 2000; Webster-Stratton, 1996).

A prerequisite for the interpretation of gender differences is that the same construct has been measured for both boys and girls (Mellenbergh, 1989; Meredith, 1993). Sex-differences have been found in the measurement of teacher-rated physical aggression in middle childhood (Miller, Vaillancourt, & Boyle, 2009). The item 'Gets into many fights' showed poorer factor loading on physical aggression for girls than boys, which indicated that an equal increase in fighting behavior (i.e., observed score) was associated with a larger increase in physical aggression (i.e., trait) for girls than boys. Thus it appeared that aggressive girls fight less often than aggressive boys, probably because physical fighting is more typical of boys in general. Measurement invariance denotes that all parameters of the factor model are invariant including also the intercepts, as intercept differences across groups may yield systematically lower or higher scores for identifiable groups. Sex-differences in intercepts have not yet been tested in measures for young children. We therefore sought to extend prior work by examining invariance across sex in both factor loadings and intercepts, and by using a sample of younger children.

The first goal of the present study was to identify a factor representing physical aggression not confounded with other antisocial behaviors. We examined the factor structure of the Preschool Behavior Questionnaire (PBQ; Behar, 1977) in two independent samples of

kindergartners. The PBQ was chosen because it was explicitly designed for use with young children, and is widely used in international research including school-based studies (e.g., Drugli & Larsson, 2006; Goossens, Bokhorst, Bruinsma, & Van Boxtel, 2002; Silver et al., 2005; Tremblay, Desmarais-Gervais, Gagnon, & Charlebois, 1987). In addition, the checklist contains a diversity of antisocial behavior descriptions but is still short and is easy to complete for teachers. Based on prior research, four items were selected representing physical aggression (see Table 2). The four other items represented nonaggressive forms of antisocial behavior. The hypothesized two-factor model was tested against a one-factor model for both boys and girls, and cross-validated in an independent sample. To obtain preliminary evidence for the model, associations with internalizing behavior were studied. We presumed that internalizing behavior would be negatively correlated with physical aggression but positively with nonaggressive antisocial behavior. Because inhibition and social withdrawal present different risks for later delinquency in boys (Kerr, Tremblay, Pagani, & Vitaro, 1997), we studied different subtypes of internalizing behavior.

The second goal was to investigate factorial invariance of the model across sex. We expected the item 'Fights' to have a higher factor loading on physical fighting for boys. Latent mean comparisons were conducted taking into account noninvariance. Boys were expected to be more physically aggressive compared to girls. No gender differences were anticipated for nonaggressive antisocial behavior.

Method

Participants. Two independent samples were included from regular elementary schools from both rural and non-rural areas of the Netherlands. In Sample A, 19 kindergarten teachers (all female) reported on preferably all children in their class (N=487; 246 girls). Children's mean age was 65.2 months (*SD*=7.7). Sample B consisted of 84 teachers (four men) and 1557

children (762 girls) with a mean age of 67.8 months (*SD*=8.3). Written informed consent was obtained from parents.

Measures. Teachers completed a Dutch version of the Preschool Behavior Questionnaire (PBQ; Behar, 1977; Goossens, Dekker, Bruinsma, & De Ruyter, 2000). This checklist contains age-appropriate descriptions of behavior problems of young children that are rated on a 4-point Likert scale (absolutely not characteristic-very characteristic). The Externalizing scale contains 14 items, including 8 items of antisocial behavior. High internal consistency ($\alpha \geq .91$), test-retest stability ($rs \geq .84$), and interrater agreement ($\alpha = .91$) have been found for the Externalizing scale in community and clinical samples. The validity has been supported by concurrent and predictive associations with parallel teacher-rated adjustment measures (concurrent *r*s= .41-.78), peer-rated aggression (*r*=.30), and children's sociometric status (Goossens et al., 2000; Goossens et al., 2002). The a-priori scale Physical Aggression (PA) included four items (see Table 2). The remaining four items were labelled Nonaggressive Antisocial Behavior (NAB).

Coplan and Armer (2007) discussed the importance of subtypes of social withdrawal in young children. Factor analyses on a set of internalizing behavior items have yielded three subscales (Thijs, Koomen, De Jong, Van der Leij, & Van Leeuwen, 2004): Social Inhibition refers to shyness or social anxiety (5 items: e.g., 'Shy or timid'). Solitary Behavior reflects social withdrawal out of unsociability or social disinterest (4 items: e.g., 'Somewhat on his/her own'), and Emotional Dysregulation reflects negative emotionality (5 items: e.g., 'Easily upset'). Internal consistencies were adequate (α = .81-.87), and the validity of the subscales was supported by unique associations with teacher interviews (Thijs et al., 2004).

Statistical analyses. Confirmatory factor analyses (CFA) were conducted using structural equation modeling (SEM) within the Mplus program. Given the hierarchical structure and the non-normal distribution of the data, the Yuan-Bentler chi-square statistic was used to

evaluate overall model fit (Muthén & Muthén, 1998-2004). Because the chi-square is sensitive to sample size, additional fit indexes were examined as well. The model fit was considered satisfactory when CFI≥.95, SRMR≤.08 and RMSEA≤.06 (Hu & Bentler, 1999). Analyses were performed under the assumption of Missing At Random. Robust maximum likelihood estimation was used, taking into account the dependency of the data due to the cluster sampling.

Measurement invariance of the constructs across gender was examined by means of factorial invariance (Meredith, 1993). Weak factorial invariance refers to equality of factor loadings, whereas strong factorial invariance refers to the invariance of both factor loadings and intercepts. Invariance for all parameters frequently does not hold and for that reason partial factorial invariance is generally considered sufficient for valid group comparisons, with at least one invariant item next to the reference indicator (Byrne, Shavelson, & Muthen, 1989). A series of multi-group analyses were conducted, starting with an unconstrained model in which factor loadings and intercepts are estimated separately in each group. When a parameter can be constrained to be equal across gender without a significant decrease in model fit, the parameter is considered invariant. When a parameter appeared as non-invariant the constraint was relaxed, allowing the estimates to take on separate values for boys and girls. First, equality constraints were imposed on all factor loadings. When the model fit increased significantly, Modification Indices (MI) were inspected and constraints were tested successively to detect the source of non-invariance. Second, equality constraints were imposed on the intercepts. Additionally, equality of structural parameters as factor variances and factor means was examined. Since constraints were evaluated multiple times, the alpha level was adjusted downward to 0.01 to consider chance capitalization. Furthermore, because the Yuan-Bentler chi-square cannot be used for difference testing of nested models, the Satorra-Bentler scaled chi-square difference test was used (Satorra & Bentler, 2001).

Results

Model confirmation and validation. The two-factor model was tested against a one-factor model in Sample A. Model fit indices are presented in Table 1, and standardized factor loadings in Table 2. Based on the MI, one residual correlation was allowed between items of the same factor (i.e., items 5 and 6). The one-factor model showed a poor fit for boys. The Satorra-Bentler scaled chi-square difference test showed a significantly better fit for the two-factor than the one-factor model ($\Delta\chi^2$=13.303, Δdf=1, p<.001). For girls, the two-factor model also fitted the data significantly better ($\Delta\chi^2$=14.957, Δdf=1, p<.001). The two-factor model was validated in Sample B and showed good fit for both boys and girls.

Factorial invariance and gender differences. Factorial invariance across gender was explored for PA and NAB in separate models in Sample B. Because the choice of the reference indicator could have influenced the results, the procedure was conducted with different reference items but virtually similar results were found. First, factor loadings were constrained to be equal across gender. The model fit for PA decreased significantly ($\Delta\chi^2$=18.520, Δdf=3, p<.01). The item 'Bullies' showed a higher factor loading for girls ($\Delta\chi^2$=126.967, Δdf=1, p<.01), whereas the item 'Fights' was found to have a higher factor loading for boys ($\Delta\chi^2$=7.407, Δdf=1, p<.01). Second, constraints across gender were imposed on the intercepts. As a result, the model fit decreased significantly ($\Delta\chi^2$=19.108, Δdf=3, p<.01). The item 'Bullies' showed a higher intercept for girls than boys ($\Delta\chi^2$=51.229, Δdf=1, p<.01). Finally, parameters of the structural model were examined. Factor variances ($\Delta\chi^2$=12.744, Δdf=1, p<.01) and factor means ($\Delta\chi^2$=30.123, Δdf=1, p<.01) were found to be unequal across gender: More variance and a higher factor mean were observed for boys (Cohen's d = *.41)*. Factor means and standard deviations for both sexes are presented in Table 2. The fit between the final model and the unconstrained model did not differ significantly ($\Delta\chi^2$=4.560, Δdf=3, p>.05).

For NAB, imposing equality constraints on the factor loadings did not significantly lower the model fit ($\Delta\chi^2$=2.698, Δdf=3, p>.05). But when constraints were added on the intercepts, the model fit decreased significantly ($\Delta\chi^2$=18.296, Δdf=3, p<.01). The item 'Inconsiderate' showed a higher intercept for boys than girls ($\Delta\chi^2$=7.015, Δdf=1, p<.01). Factor variances appeared invariant across gender ($\Delta\chi^2$=0.105, Δdf=1, p>.05). A somewhat higher factor mean emerged for boys ($\Delta\chi^2$=10.831, Δdf=1, p<.01; Cohen's d = *.20*). The fit between the final model and the unconstrained model did not differ significantly ($\Delta\chi^2$=7.873, Δdf=6, p>.05).

Furthermore, we were interested in gender mean differences in either type of antisocial behavior when corrected for the other type. This was accomplished by conducting analyses largely similar to conventional covariance analyses. The main difference is that the covariate in the present model is a latent variable instead of a measured variable. For PA, all eight items were allowed to load on one factor reflecting common variance, while the items reflecting PA were allowed to double load on a second factor reflecting unique variance. The correlation between the two factors was fixed to zero. Mean differences on the second factor were tested. The reverse was done to estimate NAB while controlling for PA. When NAB was controlled for, boys were more physically aggressive (p<.01; Cohen's d=.51). In contrast, girls appeared more nonaggressive antisocial than boys when PA was controlled for (p<.01; Cohen's d=.21).

Internalizing behavior. The model was extended with three factors representing Social Inhibition, Solitary Behavior, and Emotional Dysregulation. Parameters were constrained across gender. As a consequence the goodness of fit decreased but was still satisfactory: χ^2 (463, N=1557)= 1138.302, p<.001; RMSEA=.043; SRMR=.051; CFI=.947. PA was negatively related to Social Inhibition and for boys also positively to Emotional Dysregulation (see Table 2). NAB was positively related to Solitary Behavior and Emotional Dysregulation for both sexes.

This is an Author's Accepted Manuscript of an article published in Journal of Psychoeducational Assessment, 2010, 28, 129-138. © SAGE Publications, the final version is available online at http://dx.doi.org/10.1177/0734282909340236

Discussion

The present study established evidence for physical aggression as a distinct construct from nonaggressive forms of antisocial behavior in young children. Furthermore, results revealed noninvariance across gender in the measurement of physical aggression.

Our first goal was to provide evidence for the discrimination of physical aggression from other troublesome antisocial behaviors using CFA. A two-factor model of physical aggression and nonaggressive antisocial behavior showed a significant better fit than a one-factor model for both boys and girls, and could be validated in an independent sample. This finding is critically important as physical aggression appears a distinct risk factor for juvenile violence (e.g., Broidy et al., 2003; Loeber & Hay, 1997; Nagin & Tremblay, 1999; NICHD Early Child Care Research Network, 2004).

Differential associations with internalizing behavior provided support for the validity of the model. Largely consistent with prior studies (Crick & Grotpeter, 1995; Kazdin, 1992), for both sexes, mainly negative associations were found with physical aggression. In contrast, positive linkages emerged with nonaggressive antisocial behavior. Also, whereas physical aggression was mainly associated with social inhibition, nonaggressive antisocial behavior was related to solitary behavior and emotional dysregulation. These patterns substantiate the distinctiveness of the two categories.

The second goal was to detect factorial invariance in teacher-reported physical aggression. Unlike prior research, we tested differences in both factor loadings and intercepts. As expected, the item ‘Fights’ showed a higher factor loading on physical aggression for boys, which indicated that an equal increase in fighting reflected a larger increase in physical aggression for girls than boys. Most likely, fighting is viewed as more typical of boys, so that when girls show gender-atypical behavior and frequently engage in fights this is especially indicative of physical aggression. The opposite was seen for bullying. Both a higher factor

loading and intercept were found for girls. This suggests that girls bully more than boys given the same level of physical aggression, *and* that this difference becomes larger when physical aggression increases. Since bullying has a verbal connotation in the Dutch language, it may be that teachers interpreted the item 'Bullies' as reflecting largely verbal acts. Girls may bully more than boys because they are more competent to express themselves verbally than boys (e.g., Galsworthy, Dionne, Dale, & Plomin, 2000). Regarding nonaggressive antisocial behavior, we found an unequal intercept for the item 'Inconsiderate'. Thus boys appeared more inconsiderate of others than girls, given the same level of nonaggressive antisocial behavior. Girls could be more considerate of others in general as a result of socialization processes (Keenan & Shaw, 1997).

Though the latent means reflected somewhat different meanings for boys and girls, sufficient invariance across sex was found for both scales to allow for valid group comparisons (Byrne et al., 1989). As expected, boys displayed more physically aggression than girls but not clearly more nonaggressive antisocial behavior (cf. Maughan et al., 2000; Webster-Stratton, 1996). In sum, gender differences in mean levels and measurement were found for physical aggression, whereas mainly cross-sex similarity was established for nonaggressive antisocial behavior. This supports the view that physical aggression is a gender-specific manifestation of antisocial behavior that needs to be distinguished from less gender-specific forms.

For practitioners, this study draws attention to the heterogeneity of children's antisocial behavior and the need to use a pure measure of physical aggression that is not confounded by other antisocial behaviors in order to accurately identify children in need of preventive intervention (e.g., Broidy et al., 2003; Joussemet et al., 2008; Tremblay, 2000). The results support the use of the PBQ as a screening measure of early physical aggression. However, given the finding of unequal measurement across gender, scores do not have entirely similar

meanings for boys and girls. Physically aggressive girls seem to use more verbal threats, whereas physically aggressive boys engage more in fights. It is gender-a-typical behavior that may be especially a reason for concern. To further evaluate the impact of non-invariance for screening purposes, future research could adopt a similar approach as Millsap and Kwok (2004). Though this study draws attention to early physical aggression, children's aggressive behavior should be assessed and interpreted in the context of other key variables such as family adversity and for example comorbid internalizing problems (e.g., Kerr et al., 1997).

Several qualifications should be considered. Factorial invariance was studied largely explorative and replication is warranted. Furthermore, the results were limited to community samples and teacher reports. However, teachers play a vital role in the detection of at-risk children, and have shown to provide reliable reports (Goossens et al., 2000; Konold & Pianta, 2007).

In conclusion, we extended evidence for physical aggression as a distinct construct from nonaggressive antisocial behavior to kindergarten boys and girls. The discrimination was tested and validated using CFA, and further evidenced by differential associations with internalizing behavior and different outcomes with respect to gender differences. This is considered highly important since early physical aggression is a key variable in the prediction of chronic aggression. Furthermore, the results highlighted the necessity for researchers to consider measurement bias with respect to gender. Moreover, though non-invariance may be considered a limitation, it also advances understanding of the nature of physical aggression that is measured in boys and girls.

Tables

Table 1

Model fit indices of factor models for boys and girls in Sample A and Sample B

Model	χ^2	RMSEA	SRMR	CFI
Sample A				
Boys				
1-factor model	$\chi^2(19, N=241)=60.982$, $p<.001$	.096	.070	.915
2-factor model	$\chi^2(18, N=241)=29.256$, $p=.045$	.051	.039	.977
Girls				
1-factor model	$\chi^2(19, N=246)=73.692$, $p<.001$	.108	.059	.876
2-factor model	$\chi^2(18, N=246)=26.434$, $p<.001$	.044	.036	.981
Validation 2-factor model Sample B				
Boys	$\chi^2(18, N=795)=42.381$, $p<.01$	.041	.025	.984
Girls	$\chi^2(18, N=762)=48.247$, $p<.01$	.047	.033	.978

This is an Author's Accepted Manuscript of an article published in Journal of Psychoeducational Assessment, 2010, 28, 129-138. © SAGE Publications, the final version is available online at http://dx.doi.org/10.1177/0734282909340236

Table 2

Standardized CFA solution for boys (n=241)/ girls (n=246) for Sample A

Items	Physical Aggression	Nonaggressive Antisocial
1.Kicks, hits	.82/.75	-
2.Bullies	.83/.85*	-
3.Fights	.82/.79*	-
4.Destructive	.65/.46	-
5.Does not share	-	.64/.67
6.Inconsiderate	-	.78/.76
7.Sneaky	-	.71/.79
8.Blames others	-	.72/.87
Factor correlations		
Nonaggressive Antisocial	.75/.76	

*Note: *=Unstandardized factor loading is non-invariant across gender in Sample B*

Table 3

Descriptive statistics and correlations with internalizing behavior in Sample B

Factors	*M*	*SD*	Social Inhibition	Solitary Behavior	Emotional Dysregulation
Boys (*n*=795)					
Physical Aggression	1.17	.29	-.13***	.00	.11*
Nonaggressive Antisocial	1.29	.39	.00	.24**	.25***
Girls (*n*=762)					
Physical Aggression	1.07	.18	-.19***	.00	.00
Nonaggressive Antisocial	1.21	.39	.00	.21**	.14***

*Note: *p<.05, ** p<.01, *** p<.001*

References

Achenbach, T. M., Conners, C. K., Quay, H. C., Verhulst, F. C., & Howell, C. T. (1989). Replication of empirically derived syndromes as a basis for taxonomy of child/adolescent psychopathology. *Journal of Abnormal Child Psychology, 17*, 299-323.

Behar, L. B. (1977). The preschool behavior questionnaire. *Journal of Abnormal Child Psychology, 5*, 265-275.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39*, 222-245.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development, 79*, 1185-1229.

Coplan, R. J., & Armer, M. (2007). A "multitude" of solitude: A closer look at social withdrawal and nonsocial play in early childhood. *Child Development Perspectives, 1*, 26-32.

Cote, S., Vaillancourt, T., LeBlanc, J., Nagin, D., & Tremblay, R. (2006). The development of physical aggression from toddlerhood to pre-adolescence: A nation wide longitudinal study of canadian children. *Journal of Abnormal Child Psychology, 34*, 68-82.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66*, 710-722.

Drugli, M., & Larsson, B. (2006). Children aged 4-8 years treated with parent training and child therapy because of conduct problems: Generalisation effects to day-care and school settings. *European Child & Adolescent Psychiatry, 15*, 392-399.

Eley, T. C., Lichtenstein, P., & Stevenson, J. (1999). Sex differences in the etiology of aggressive and nonaggressive antisocial behavior: Results from two twin studies. *Child Development, 70*, 155-168.

Frick, P. J., Lahey, B. B., Loeber, R., Tannenbaum, L., Van Horn, Y., Christ, M. A. G., et al. (1993). Oppositional defiant disorder and conduct disorder: A meta-analytic review of factor analyses and cross-validation in a clinic sample. *Clinical Psychology Review, 13*, 319-340.

Galsworthy, M. J., Dionne, G., Dale, P. S., & Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. *Developmental Science, 3*, 206-215.

Goossens, F. A., Bokhorst, K., Bruinsma, C., & Van Boxtel, H. W. (2002). Judgments of aggressive, withdrawn and prosocial behavior: Perceived control, anger, pity and sympathy in young dutch children. *Journal of School Psychology, 40*, 309-327.

Goossens, F. A., Dekker, P., Bruinsma, C., & De Ruyter, P. A. (2000). *De gedrags vragenlijst voor peuters en kleuters: Factor structuur, betrouwbaarheid en validiteit. [behavior questionnaire for 2- to 6-year-olds: Factor structure, reliability, and validity]*. Amsterdam: Vrije Universiteit.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Joussemet, M., Frank Vitaro, F., Barker, E. D., Côté, S., Nagin, D. S., Zoccolillo, M., et al. (2008). Controlling parenting and physical aggression during elementary school. *Child Development, 79*, 411-425.

Kazdin, A. E. (1992). Overt and covert antisocial behavior: Child and family characteristics among psychiatric inpatient children. *Journal of Child and Family Studies, 1*, 3-20.

Keenan, K., & Shaw, D. (1997). Development and social influences on young girls' early problem behavior. *Psychological Bulletin, 1997*, 95-103.

Kerr, M., Tremblay, R. E., Pagani, L., & Vitaro, F. (1997). Boys' behavioral inhibition and the risk of later delinquency. *Archives of General Psychiatry, 54*, 809-816.

Konold, T. R., & Pianta, R. C. (2007). The influence of informants on ratings of children's behavioral functioning: A latent variable approach. *Journal of Psychoeducational Assessment, 25*, 222-236.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology, 48*, 371-410.

Loeber, R., & Schmaling, K. B. (1985a). Empirical evidence for overt and covert patterns of antisocial conduct problems: A metaanalysis. *Journal of Abnormal Child Psychology, 13*, 337-353.

Loeber, R., & Schmaling, K. B. (1985b). The utility of differentiating between mixed and pure forms of antisocial child behavior. *Journal of Abnormal Child Psychology, 13*, 315-335.

Maughan, B., Pickles, A., Rowe, R., Costello, E., & Angold, A. (2000). Developmental trajectories of aggressive and non-aggressive conduct problems. *Journal of Quantitative Criminology, 16*, 199-221.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127-143.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.

Miller, J. L., Vaillancourt, T., & Boyle, M. H. (2009). Examining the heterotypic continuity of aggression using teacher reports: Results from a national canadian study. *Social Development, 18*, 164-180.

Millsap, R. E., & Kwok, O. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods, 9*, 93-115.

Monuteaux, M. C., Fitzmaurice, G., Blacker, D., Buka, S. L., & Biederman, J. (2004). Specificity in the familial aggregation of overt and covert conduct disorder symptoms in a referred attention-deficit hyperactivity disorder sample. *Psychological Medicine, 34*, 1113-1127.

Moreland, A. D., & Dumas, J. E. (2008). Categorical and dimensional approaches to the measurement of disruptive behavior in the preschool years: A meta-analysis. *Clinical Psychology Review, 28*, 1059-1070.

Muthén, L. K., & Muthén, B. O. (1998-2004). *Mplus user's guide. Third edition*. Los Angeles, CA: Muthén & Muthén.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*, 1181-1196.

Nagin, D. S., & Tremblay, R. E. (2005). What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *The ANNALS of the American Academy of Political and Social Science, 602*, 82-117.

NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood. *Monographs of the Society for Research in Child Development, 96*.

Nigg, J. T., & Hinshaw, S. P. (1998). Parent personality traits and psychopathology associated with antisocial behaviors in childhood attention-deficit hyperactivity disorder. *Journal of Child Psychology and Psychiatry, 39*, 145-159.

Patrick, M. R., Snyder, J., Schrepferman, L. M., & Snyder, J. (2005). The joint contribution of early parental warmth, communication and tracking, and early child conduct problems on monitoring in late childhood. *Child Development, 76*, 999-1014.

Quay, H. C. (1987). Patterns of delinquent behavior. In H. C. Quay (Ed.), *Handbook of juvenile delinquency* (pp. 118-.138). New York: John Wiley & Sons.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514.

Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher-child relationship during the school transition. *Journal of School Psychology, 43*, 39-60.

Stanger, C., Achenbach, T. M., & Verhulst, F. C. (1997). Accelerated longitudinal comparisons of aggressive versus delinquent syndromes. *Development and Psychopathology, 9*, 43-58.

Stipek, D., & Miles, S. (2008). Effects of aggression on achievement: Does conflict with the teacher make it worse? *Child Development, 79*, 1721-1735.

Tackett, J. L., Krueger, R. F., Sawyer, M. G., & Graetz, B. W. (2003). Subfactors of dsm-iv conduct disorder: Evidence and connections with syndromes from the child behavior checklist. *Journal of Abnormal Child Psychology, 31*, 647-654.

Thijs, J. T., Koomen, H. M. Y., De Jong, P. F., Van der Leij, A., & Van Leeuwen, M. G. P. (2004). Internalizing behaviors among kindergarten children: Measuring dimensions

of social withdrawal with a checklist. *Journal of Clinical Child and Adolescent Psychology, 33*, 802-812.

Tremblay, R. E. (2000). The development of aggressive behaviour during childhood: What have we learned in the past century? *International Journal of Behavioral Development, 24*, 129-141.

Tremblay, R. E. (2004). Decade of behavior distinguished lecture: Development of physical aggression during infancy. *Infant Mental Health Journal, 25*, 399-407.

Tremblay, R. E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P. (1987). The preschool behaviour questionnaire: Stability of its factor structure between cultures, sexes, ages and socioeconomic classes. *International Journal of Behavioral Development, 10*, 467-484.

Vaillancourt, T., Brendgen, M., Boivin, M., & Tremblay, R. E. (2003). A longitudinal confirmatory factor analysis of indirect and physical aggression: Evidence of two factors over time? *Child Development, 74*, 1628-1638.

Webster-Stratton, C. (1996). Early-onset conduct problems: Does gender make a difference? *Journal of Consulting and Clinical Psychology, 64*, 540-551.